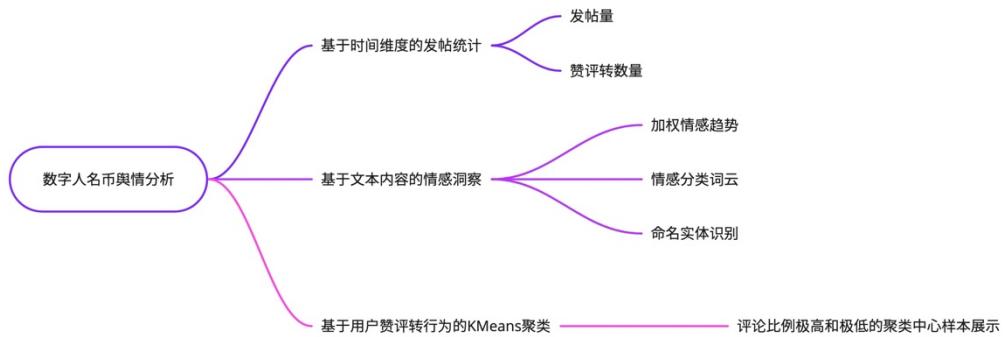


Overview:



- **发帖量从 2021 年 4 月至 7 月呈上升趋势，平均每个帖子的转赞评数量也符合发帖趋势。中文帖文中的转赞评数量相对于发帖量展现出一定滞后性，但这一现象在英文帖文中没有得到验证**
——滞后性取决于新闻讯息的传播速度和渠道。英文中规律的缺失归因于不平衡的中英数据比例 4:1
- **整体舆情倾向呈正面，自 2021 年 3 月上涨于 6 月回跌，之后持续增长。中英评论的情感峰值，在 2021 年 5 月和 7 月，出现错峰现象**
——情感分值的错峰意味着同一新闻国内外群众不同关注度、反响，以及媒体不同程度的宣扬和打压。5 月“五五购物节”的试点活动和支付宝推出数字人民币模块等服务在国内掀起浪潮。7 月中国人民银行向世界发布的《中国数字人民币的研发进展》白皮书在国外引起激烈讨论
- **正面帖文主要跟金融商圈相关。负面帖文主要跟日常消费和利润相关，有近 50% 上榜的负面关键词是繁体字体**
——负面帖文发布者大多来自于港澳台地区
- **中国电信作为运营体系下的发行商，提及次数仅次于中国银行。除了京津冀、长三角、粤港澳大湾区之外，地点识别中出现了成都、三亚、长春、郑州等地。境外地点中日本、印度、伊朗、新加坡和马来西亚等国也频频上榜**
——国内中西部地区是自 2021 年 2 月扩张的第二批试点场景。此外，海南预计 11 月实施的全省试点，截止 9 月为止，提及次数在实体统计中排名第 10。境外，大多国家支持数字人民币跨境支付，与中国达成紧密合作关系；日本和印度央行效仿中国开展自己国家的数字货币计划；秉持着“不承认不禁止虚拟货币 公众决定未来”的马来西亚，也以第 15 的排名上了实体统计榜单
- **基于用户的赞评转行为训练聚类模型，展示评论比例极高和极低的聚类中心附近样本，发现评论的内容属于新闻报道，文本的情感分类 75% 为正面**
——只评论不点赞转发的用户和只点赞转发不评论的用户，关注的大多数描述客观事实的新闻

1 基于时间维度的发帖量概观

1.1 总体发帖量及点赞评论转发数统计

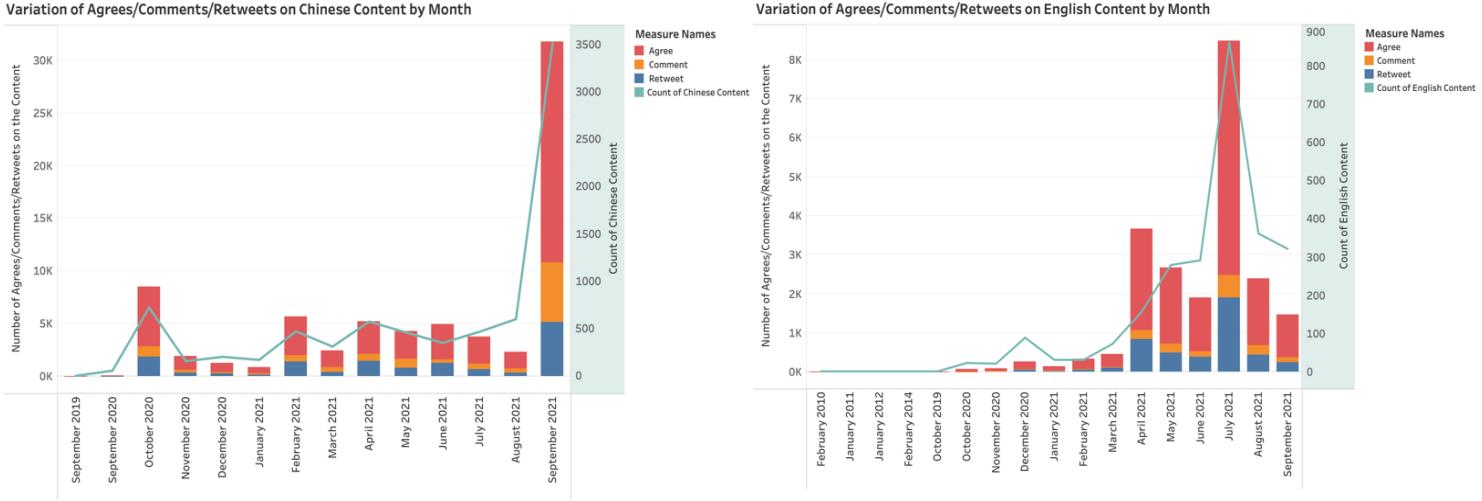


Figure 1(左)中文原文帖文发帖量(折线)、 赞评转累计数量随月份变化(柱状图)。

(右)英文原文帖文发帖量(折线)、 赞评转累计数量随月份变化(柱状图)。

爬取 2010 年 1 月至 2021 年 9 月，总计爬取到原帖数据 11676 条，筛选中英文本进行分析。

根据左图中文发帖统计，中文帖文总量在 2021 年 9 月出现暴涨，其次在 2020 年 10 月和次年 4 月。2020 年 10 月和次年 4 月的上涨，分别跟深圳和广州接连开展的数字人民币试点活动相对应。2021 年 9 月的发帖峰值可被解释为收集的爬虫数据聚焦于 9 月。

根据右图英文发帖统计，英文发帖总量在 2021 年 7 月出现暴涨，其次在 2021 年 4 月和 5 月。与之对应的时事为 2021 年 7 月举办的世界人工智能大会，和中国人民银行发布《中国数字人民币的研发进展》白皮书。

中英发帖共同规律可被总结为：2021 年 4 月发帖量开始上升，尤其是 7 月后中英文帖文总量均出现急剧上升。整体而言，赞转评数量随发帖量增加而增加。

1.2 基于用户和基于帖文的点赞评论转发数统计

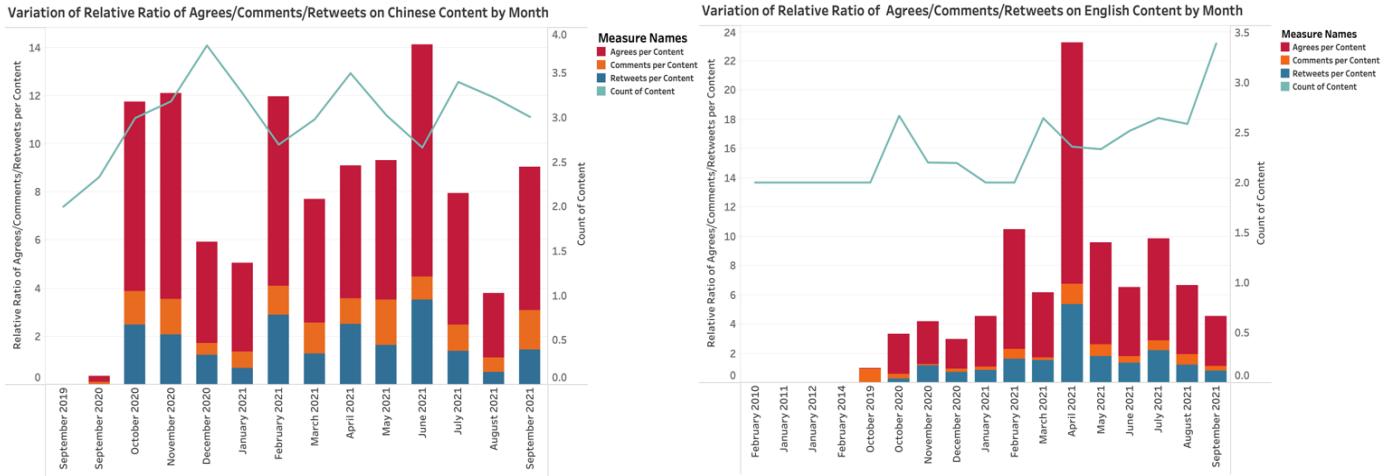


Figure 2 (左)中文帖文的人均发帖量变化(折线), 帖均转评赞数量变化(柱状图)

(右)英文帖文的人均发帖量变化(折线), 帖均转评赞数量变化(柱状图)

人均发帖量(折线)和帖均转评赞数量(柱状图)的相对趋势, 可以反映关注度的变化。中文帖文的平均关注度极值出现于 2021 年 4 月, 在 2020 年 12 月和次年 8 月观察到相反趋势。随人均发帖量上涨, 帖均转评赞数量反而下降。对于英文帖文, 在 2021 年 2 月和 6 月也出现类似的相反趋势。相悖趋势反映的可能是发布原帖的用户的关注度突增, 或是赞评转用户的关注度下降。

整体观察折线和柱状图的变化趋势, 可以观察到转评赞数量的趋势随发帖量的变化有一定滞后性, 这在中文帖文中尤其明显, 发帖量的三个峰值(折线)在转评赞数量(柱状图)的两个月之后得到体现。但在英文帖文中这种滞后性并没有观察到, 考虑到英文数据量相对较少, 采取分析的中英比例为 4:1, 所以英文帖文的滞后特点还有待考证。

2. 基于文本内容的情感洞察

2.1 加权情感分析

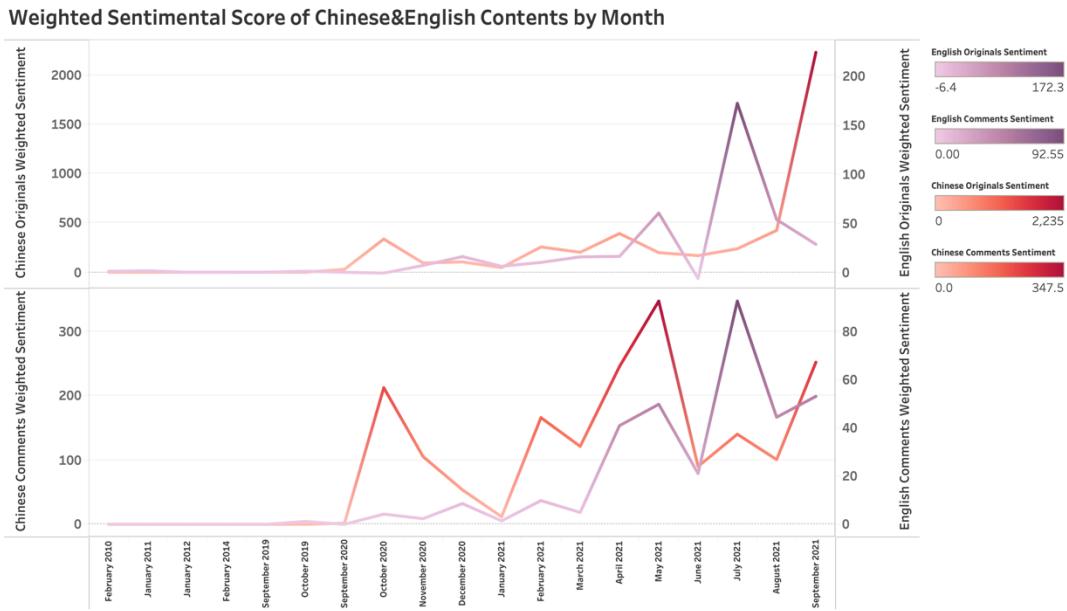


Figure 3 中英原帖和评论内容的加权情感打分，上轴为中英原帖的情感变化，下轴为中英评论的情感变化

整体而言，加权情感打分呈正面且持续上升趋势，可观察到评论(下轴)比原贴(上轴)的情感波动幅度更剧烈。由评论的加权打分可观察到，中英评论的情感均在 2021 年 1 月和 6 月经历显著下滑至当年低谷。与之对应的时事有于北京、成都、苏州等地的试点。

中文原帖的情感波动折线较为平稳，相比之下，中文评论折线在 2020 年 10 月和次年 5 月达到峰值。英文原帖和评论的情感波动呈现一致的波动规律，从 2021 年 3 月逐步上升，并于 7 月达到峰值。英文帖文中相对不明显的情感趋势需考虑到数据量本身的缺少，不足以反映足够明显的波峰波谷。

对比中英评论情感的峰值，可观察到在 2021 年 5 月和 7 月，出现错峰现象：5 月国内舆论情感出现高峰但国外并无对应趋势。7 月国外舆论情感出现高峰，而国内反响相对低迷。所造成的原因可能是，对于同一新闻国内外不同程度的宣扬和打压，低峰可能来自模糊掩饰的报道，高峰来自信息的刻意推广，同时也有可能是激烈的舆论反馈。与之关联的事件有 5 月国内的“五五购物节”、支付宝首次开通数字人民币支付；7 月人民银行发布的白皮书于海外引起的热烈讨论。

2.2 中英文词频词云



Figure 4 正面帖文关键词 正面帖文相悖情感词语

在分类为正面的中文帖文中，top15 关键词普遍跟金融商圈相关：丽泽商务区、普惠、金融体系等。在这些正面文本中，被识别为相悖情感的词语包括“网络系统”、“信息技术”、“攻陷”、“有待”；人名提及“冯远征”和中国银行原行长‘李礼辉’。

在分类为正面的英文帖文中，出现较多信用卡(mastercard engage&draw)和加密货币(cryptos)等相关词汇。文本中出现比例较高的负面情感的词语有 criticize 和 liability。



Figure 5 负面帖文关键词 负面帖文相悖情感词语

在分类为负面的中文帖文中，top15 关键词普遍跟日常消费和利润相关：餐厅、利润率、持仓等。同时发现，情感负面帖文的关键词有一半是繁体字体。在这些负面文本中，整体提及正面情感的词语普遍较少，繁体字体的“现金”和简体字体的“省去”出现次数最多。

在分类为负面的英文帖文中，出现较多地名(salvador, york, dorsey)、薪水(salaries)和死亡率等关键词。文本中比例较高的相悖情感词语有警示(alert)和解决(solved)。

2.3 命名实体识别

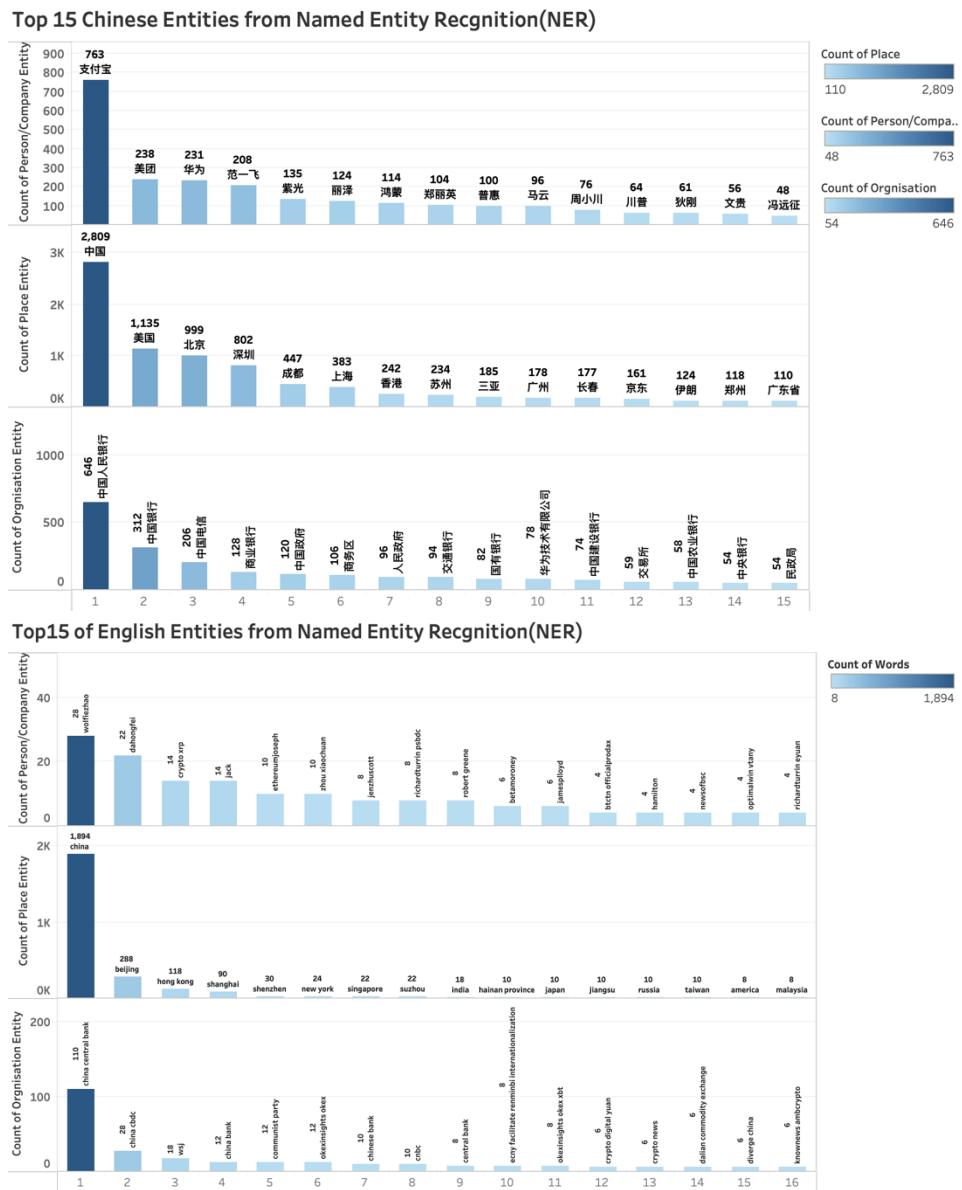


Figure 6 中英文人名/企业、地名、机构团体实体识别

命名实体识别中，人名中提及最多是中国人民银行副行长范一飞、原行长周小川和马云。

媒体名称有区块链媒体主编赵卫 wolfiezhao、区块链顾问达鸿飞、以太坊联合创始人 Joseph、卡内基国际和平基金会常驻学者 Robert Greene。

地区实体主要集中于珠三角、长三角地区，其他像成都、三亚、长春、郑州等地分别作为第二批的扩张试点场景被提及。其中海南预计 11 月实施的全省试点，根据数据截止 9 月，提及次数在英文文本中排名第 10。其次，在 top10-15 中，境外像美国、日本、伊朗、新加坡、印度等国也被提及较多，日本和印度采取的行动是效仿中国开创自己国家的数字货币。伊朗、新加坡、俄罗斯和马来西亚选择支持数字人民币跨境交易，均与中国央行展开紧密合作。

机构团体提及最多的是中国人民银行、政府和共产党。企业标识中“支付宝”作为运行体系的流通层，被提及次数是中国电信和商业银行的近4倍。美团、华为、京东作为应用场景被提及；紫光国微和鸿蒙作为技术支持被提及。

3. 基于点赞评论转发的聚类模型探索

3.1 使用簇内平方 SSE 和轮廓系数确认的聚类中心数 k 值检验

基于余弦距离对原帖帖文点赞、评论、转发数训练 KMeans 聚类模型

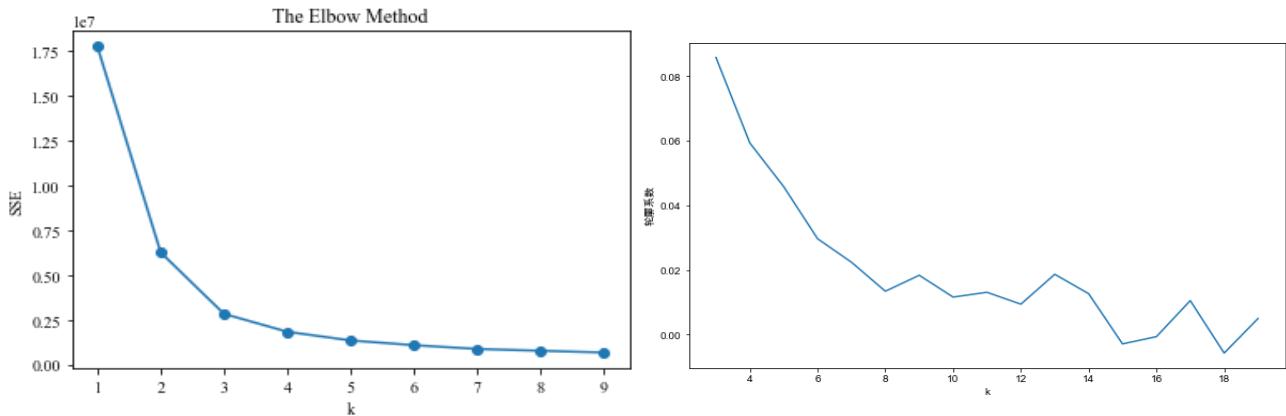


Figure 7 使用簇内平方 SSE(左)和轮廓系数(右)

SSE 利用计算每个 K 值对应的簇内点到中心的距离误差平方和，SSE 的值越小，代表聚类效果越好。观察 SSE 曲率(左)，由肘部法得到当 k 取 3 或 4 时，出现最大转折点，降低幅度越大，聚类内相似度最高。

轮廓系数结合内聚度和分离度两种因素，评估簇内部不相似性和不同簇之间的不相似性。观察曲线(右)的畸变程度，再次验证 k 取 4 时，斜率变化最大。

3.2 聚类结果展示

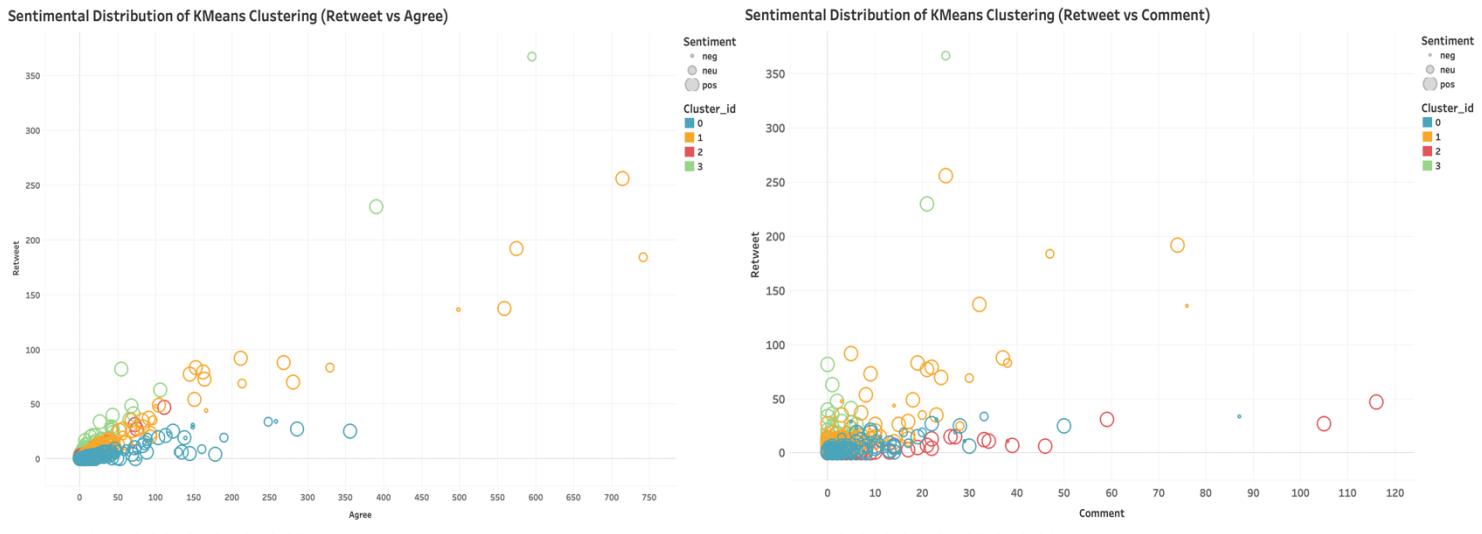


Figure 8 基于赞评转行为的 4 类用户，转发 vs 点赞，转发 vs 评论，圆圈半径代表情感打分

Agree/Comment/Retweet Proportion of Clustering Centres

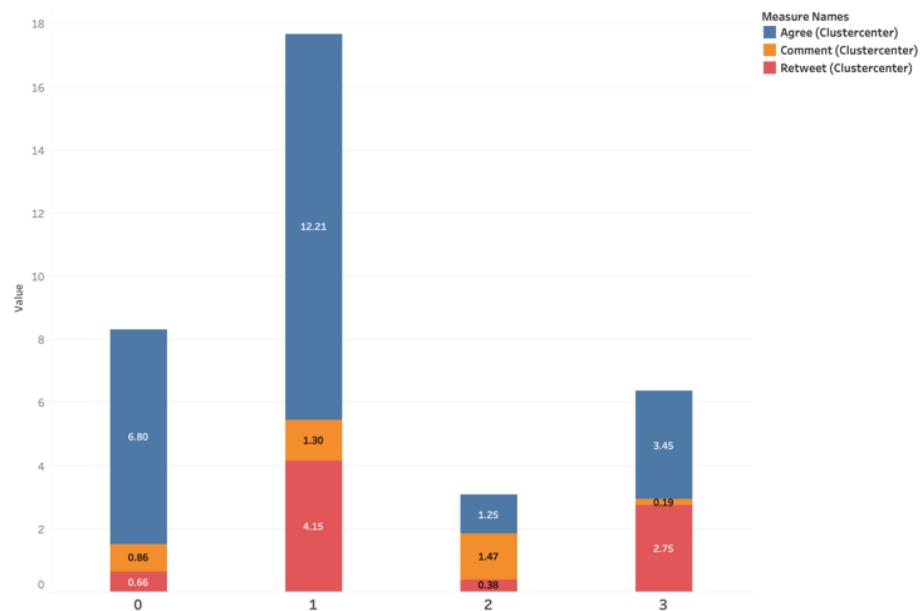


Figure 9 用户赞转评行为的聚类中心

用户行为被聚类成 4 类，如图 8 所示，这 4 类用户行为的转发数随点赞数增加而增加。在聚类中心的分布比例图上，可以观察到第 0 类用户点赞行为较高，评论和转发数相对较少且比例持平，赞评转比例 10:2:1；第 1 类用户，点赞和转发行为较高，赞评转比例为 10:1:3；第 2 类用户，评论行为是所有用户中最高，但转赞数量最低，赞评转比例为 3:4:1；第 3 类用户评论行为最少，但点赞和转发较高，赞评转比例呈 17:1:14。

从整体上观察，第 2 类和 3 类用户都是属于最具有反馈行为的群体，尤其第 2 类评论行为最高，点赞转发最低。第 3 类的评论反馈行为最少，点赞转发行为较高。

已知每一类的数据条数都在 2000 左右，平衡的样本比例再次验证了结论的可探讨性。

3.3 聚类中心附近样本和对应情感分类

cluster_id	agree	comment	retweet	content	sentiment
0	6	1	0	三分天下？数字人民币时代开启，支付宝、微信的差距有点明显	pos
	6	1	0	WSB 系列股票 AMC 暴涨 112%，狗狗币行情可期，数字人民币基于以太坊，后市将如何？	pos
	7	1	1	【李礼辉：2022 冬奥会将成为数字人民币的试金石和转折点，对此充满信心】链闻消息，据中国日报报道，中国银行前行长、中国互联网金融协会区块链研究工作组组长李礼辉表示，数字人民币试点于 2020 年 4 月启动	pos
	7	1	0	Digi-pound will come.China already has e-CNY and nearly every major currency in the world is consulting at the moment.	neu
	6	1	0	Bitcoin is over, the technological breakthrough is a national crypto currency, which the US has still not managed to do unlike China and the E-CNY ¥ which spoiler: allows to bypass the illegal US and EU sanctions.	neg
	7	1	1	#DECP or e-CNY is coming. My 1st transaction on Shenzhen street (at an interesting time.)For users and merchants, the UX is exactly like WeChat/Alipay QR payment. The major differences are in backend, b/w the commercial banks and the central bank. #centralbankdigitalcurrency	pos
1	13	1	3	数字人民币官方宣传片定调，未来核心在区块链！比特币狗狗币会大涨？数字粮票限制消费，人民币居然有保质期！数字钱包与币圈韭菜命运有何联系？	neg
	13	2	0	今天中国银行的跑来推广央行数字人民币，把我加入了 Whitelist，说过几天可以体验一下转账 0 手续费和速度，指定商户体验可以打折，还有最高 100 元的数字红包雨，我心想.....你这当然快啦，又跟区块链没啥关系，只是纸钱数字化乐而已，本质上还是干的阿里巴巴早赶去过的活儿***	pos
	12	0	0	这也算是创了历史先河吧.....新型计划经济，真不怕玩坏？ #数字人民币	neg
	13	1	3	*#Digitalyuan debuts in futures transaction*e-CNY payment charges no transaction fee and is not restricted to the operation period of the high-value payment system@psb_d @efipm @BrettKing@leimer @spirosmargaris @BetaMoroney #fintech #CashlessChina	pos
	13	2	0	CN's digital wallets are ordered to debundle payment from other financial products, the same week we learn upgrades on e-CNY.Implementation is vague but one way might be to move other financial products into a separate app.	neg
	12	0	0	This is not a drill. Cyber hornets, we must relay the risk of the Byzantine Generals to @SenShelby because he doesn't understand that 21st century war is about payments and data.(BTC + USD > e-CNY = beyond his understanding.)Time to call, email, and social media.	neg
2(评论比例最高，赞转最少)	1	2	0	央行前行长周小川：支付系统现代化是隔一段时间迈一个台阶“要有比较大的台阶，大家才有足够的动力去迈这个台阶。就像手机更新换代，有的厂商几个月就发表一个版本。特别小的台阶，微小的改进，就要换系统的话，有可能动力不一定足。”#DCEP #数字人民币	neg
	1	2	0	【深圳市启动数字人民币在公共交通绿色出行领域试点工作，市民可用数字人民币乘公交地铁】链闻消息，据深圳特区报消息，深圳市交通运输局联合中国人民银行、农业银行、建设银行和深圳通有限公司，正式启动数字人民币在公共交通绿色出行领域试	pos
	1	2	0	王毅敦促 G20 大国勿将“五桩罪”殃及中国；中美拔河抢普金，布林肯：美国还是想与普京站一边；中俄联手日本紧张，称保护台湾民主国家被中国狠批；日本：中国数字人民币不可忽视；北京助力“中国制造，亚马逊销售”	neg
	1	2	0	China flocks to Dogecoin-like Shiba, e-CNY for kindness: Blockheads	pos
	1	2	0	You might not agree on how they see BTC (I don't) but if you don't pay attention to this, then you won't get "it". Also: EU and US *will* have CBDCs as well, e-CNY will be to most fair and inclusive (sic!) one though.	pos
	2	1	0	China's digital currency (e-CNY) will track who uses it and on what items and activities, and when. So either be prepared to live without a wallet (if one values privacy) or stop accepting bribes. The latter effect would be good for Bangladesh (when we'll have e-BDT)!	pos
3(评论比例最低，赞转较多)	3	0	3	数字人民币已经出现假冒了，纸质钞票的核心技术还没折腾明白，就去搞这些花里胡哨的。真怕那一天被人破解了，又搞成金圆券。	pos
	4	0	2	上海：建设 100 个支持数字人民币支付的 AI 食堂 - 8 月 23 日消息，据上海市经济与信息化委员会透露，已规划在全市推广 100 个 AI 社区食堂，其中将部署 AI 识别智能结算台，实现数字人民币等多种支付方式的结算，目前初步计划到今年年底推出 10 至 20 个。（科技日报）	pos
	4	0	2	【央行：下一阶段要稳妥推进数字人民币研发试点】链闻消息，中国人民银行召开 2021 年货币金银和安全保卫工作电视会议，会议强调，下一阶段，要切实保障现金供应，稳妥推进数字人民币研发试点，持续深入做好整治拒收现金工作，不断完善	pos

3	3	0	3	#China's e-CNY White Paper Criticizes Bitcoin; Reveals #CBDC Project Uses Smart Contract BY: Anthonia Isichei https://t.co/n29eFIAVO #CentralBankDigitalCurrency #DigitalYuan #ECNY #PBOC	pos
	3	0	3	China's e-CNY digital currency could have a significant impact on China's domestic economy and on geopolitics. My #CarnegieAsia colleague Robby Greene explains why:	pos
	3	0	3	Central Bank Digital Currencies (CBDCs) are here and rising at an accelerated pace. 24 million e-cny (e-chinese yuan) wallets funneling already 34.5 billion yuan. Digitalization of corporate and personal finance is not a luxury, it is a must for corpor...	neg